PB-245 052

SOURCE TEST DATA SYSTEM (SOTDAT)

TRW/Environmental Services
Vienna, Virginia

July 1975

DISTRIBUTED BY:

NTIS

National Technical Information Service
U. S. DEPARTMENT OF COMMERCE

TECHNICAL REPORT DATA
(Please read Instructions on the reverse before completing)

1. REPORT NO.	2.	3. RECIPIENT'S ACCESSION NO.
EPA-450/3-75-070		

4. TITLE AND SUBTITLE	5. REPORT DATE
SOTDAT Final Report	July, 1975
	6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S)	8. PERFORMING ORGANIZATION REPORT NO.
	96005.003
9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. PROGRAM ELEMENT NO.
TRW Transportation and Environmental Engineering Operations 800 Follin Lane, SE Vienna, Virginia 22180	11. CONTRACT/GRANT NO. 68-02-1007
12. SPONSORING AGENCY NAME AND ADDRESS	13. TYPE OF REPORT AND PERIOD COVERED
U.S. Environmental Protection Agency Office of Air Quality Planning & Standards Research Triangle Park, N. C. 27711	Final Report 14. SPONSORING AGENCY CODE

15. SUPPLEMENTARY NOTES

16. ABSTRACT

Throughout the country, there is a vast amount of source test data which has been compiled in recent years. Up until now, these data have been of little use to anyone needing a large amount of data, because they are stored in so many different places and formats.

The Source Test Data (SOTDAT) System is a useful solution to that problem. The SOTDAT System permits the gathering of source test data from many places and their storage in a computer-accessible data bank in a common format. SOTDAT is designed so that each record describes, in detail, one run of a stack test. Variables included are most of those which enter into the normal stack test calculations, as well as some which will be necessary to future users of SOTDAT. Information stored in SOTDAT contains an adequate number of source parameters and concentrates heavily on data describing a specific test run. Since each SOTDAT record is keyed to a record in the National Emissions Data System (NEDS), any required source parameters are readily available from a NEDS listing. An exception to this will exist in the case where test data are coded anonymously in order to protect the confidentiality of the data. For a complete list and description of the SOTDAT variables, see the August 1973 National Air Data Branch publication "Source Test Data System (SOTDAT)" which describes in detail each data element.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
SOTDAT NEDS Emission Factors	PRICES SUBJECT TO CHANGE	

18. DISTRIBUTION STATEMENT	19. SECURITY CLASS (This Report)	21. NO. OF PAGES
Release Unlimited	Unclassified	15
	20. SECURITY CLASS (This page)	22. PRICE
	Unclassified	3.25-2.25

EPA Form 2220-1 (9-73)

268274
PB 245 052

EPA-450/3-75/070
July 1975

SOTDAT
FINAL REPORT

Reproduced by
NATIONAL TECHNICAL
INFORMATION SERVICE
US Department of Commerce
Springfield, VA. 22151

U.S. ENVIRONMENTAL PROTECTION AGENCY
Office of Air and Waste Management
Office of Air Quality Planning and Standards
Research Triangle Park, North Carolina 27711

EPA-450/3-75-070

SOTDAT
FINAL REPORT

by

TRW
Transportation and Environmental Engineering Operations
800 Follin Lane, SE
Vienna, Virginia 22180

Contract No. 68-02-1007, Task 3

EPA Project Officer: Gregory Bujewski

Prepared for

ENVIRONMENTAL PROTECTION AGENCY
Office of Air and Waste Management
Office of Air Quality Planning and Standards
Research Triangle Park, North Carolina 27711

July 1975

This report is issued by the Environmental Protection Agency to report technical data of interest to a limited number of readers. Copies are available free of charge to Federal employees, current contractors and grantees, and nonprofit organizations - as supplies permit - from the Air Pollution Technical Information Center, Environmental Protection Agency, Research Triangle Park, North Carolina 27711; or, for a fee, from the National Technical Information Service, 5285 Port Royal Road, Springfield, Virginia 22161.

This report was furnished to the Environmental Protection Agency by TRW, Transportation and Environmental Engineering Operations, Vienna, Virginia 22180, in fulfillment of Contract No. 68-02-1007. The contents of this report are reproduced herein as received from TRW, Transportation and Environmental Engineering Operations. The opinions, findings, and conclusions expressed are those of the author and not necessarily those of the Environmental Protection Agency. Mention of company or product names is not to be considered as an endorsement by the Environmental Protection Agency.

Publication No. EPA-450/3-75-070

1.0 INTRODUCTION

1.1 GENERAL DESCRIPTION OF THE SOURCE TEST DATA (SOTDAT) SYSTEM

Throughout the country, there is a vast amount of source test data which has been compiled in recent years. These data are on file in EPA offices, both in Durham and in the regions, in state and local control agency offices, with private consultants who have conducted stack tests, industrial plants where tests have been run, control equipment manufacturers, and others. Up until now, these data have been of little use to anyone needing a large amount of data, because they are stored in so many different places and formats.

The Source Test Data (SOTDAT) System is a useful solution to that problem. The SOTDAT System permits the gathering of source test data from many places and their storage in a computer-accessible data bank in a common format. SOTDAT is designed so that each record describes, in detail, one run of a stack test. Variables included are most of those which enter into the normal stack test calculations, as well as some which will be necessary to future users of SOTDAT. Information stored in SOTDAT contains an adequate number of source parameters (e.g. plant name, location, stack height, etc.) and concentrates heavily on data describing a specific test run. Since each SOTDAT record is keyed to a record in the National Emissions Data System (NEDS), any required source parameters are readily available from a NEDS listing. An exception to this will exist in the case where test data are coded anonymously in order to protect the confidentiality of the data. For a complete list and description of the SOTDAT variables, see the August 1973 National Air Data Branch publication "Source Test Data System (SOTDAT)" which describes in detail each data element.

1.2 VALUE OF SOTDAT INFORMATION

The data contained in the SOTDAT System will be useful for many purposes. The single fact which makes these data so useful is that, instead of being a mixture of measured, calculated, and estimated data as NEDS is, SOTDAT is composed entirely of measured data. This greatly increases the reliability of any deductions based on data from SOTDAT.

The most immediate use to which SOTDAT will be put is to validate and/or correct existing emission factors, and to create new ones in areas where factors have not yet been compiled. In conjunction with this use, SOTDAT could probably be used as a validity check on the NEDS system. Estimated emissions in NEDS which are grossly inconsistent with SOTDAT-generated factors could be flagged for further investigation.

Another use for data in SOTDAT is the development of accurate methods for calculating control device efficiencies based on specific operating parameters. These parameters are part of the SOTDAT data base.

A system which contains the type of basic, fundamental data that SOTDAT does, is sure to become extremely valuable in the future. Data which deal with actual (not estimated or calculated) emissions from specific pollution sources is certainly more valuable than what has been available thus far. Hopefully, the individuals charged with maintaining the SOTDAT system will be sensitive to the needs for this data, and will remain flexible enough to implement changes as they are needed.

2.0 DATA ACCUMULATION

2.1 DATA ON FILE IN THE EMISSIONS MEASUREMENT BRANCH OF EPA

Many source tests have been performed by personnel of EPA's Emissions Measurement Branch (EMB) or by EMB-obtained private contractors. Results from many of these tests are also on file in the National Air Data Branch (NADB), and were therefore available for removal from Durham. The data from these 155 reports were coded onto SOTDAT coding forms in TRW's McLean, Va. office. This effort produced 1292 completed coding forms.

Following the data validation process described in Section 3.0, the data from another 9 test reports (submitted to NADB after the original coding effort) were entered on 109 SOTDAT coding forms.

Another 26 test reports were on file in the EMB office but not in the NADB. Since these reports could not be removed, the data they contained were coded in Durham. This additional data generated 209 completed forms.

2.2 DATA ON FILE IN THE EMISSION STANDARDS AND ENGINEERING DIVISION OF EPA

The Emission Standards and Engineering Division of EPA has a file of incinerator test results located in the IRL Building in Durham. These are test reports which have been submitted to EPA in an attempt to obtain EPA certification for a specific model of an incinerator. The file contains reports on incinerators which have received certification as well as those which have been unable to meet the certification standards. Data from 68 reports were coded onto 173 coding forms during the data accumulation effort expended in this location.

2.3 SUMMARY OF RESULTS

All in all, during the project, 190 source test reports were read; data were extracted from them and coded onto 1607 SOTDAT coding forms.

The data now present in the SOTDAT system comprise a relatively good cross section of most types of industries. However, since the majority of the tests were performed to accumulate data to be used in the establishment of New Source Performance Standards, the SOTDAT data base may presently be biased toward the better controlled, or more efficient sources.

3.0 DATA VALIDATION

3.1 NEED FOR VALIDATION

After the coding effort was completed, the data were keypunched and loaded into the computer. The resulting output from the system revealed that several problems existed either in the input data, or in the computer program. It was decided that for the SOTDAT System to be a truly useful tool, it would be necessary to rectify as many of the existing problems as possible.

3.2 DATA VALIDATION EFFORT

The apparent errors noted during the initial brief examination of the computer output included missing data, erroneous values, and unexplainable printed symbols in the place of data. The approach employed to identify and correct the errors involved reviewing each output record (one per input coding form), and checking for noticeable errors of the type listed above.

Whenever suspected errors were discovered, it was necessary to determine whether an error actually existed, and if so, note it appropriately for later correction. This was accomplished by checking each entry which appeared to be wrong against the original coding form, and then, if necessary, against the data in the stack test report. True errors were noted directly on the output.

At the time this effort was taking place, it was impossible to ascertain definitive procedures for updating the data stored in the computer, so the changes were made directly on the original coding forms. This retained the maximum flexibility since either the entire form could be repunched, or just the card or cards which required correction. This resulted in approximately 300 coding forms which contained errors.

Examination of the apparent errors demonstrated many general problems with the computer program, and generated several suggestions for improving the system. They were noted during the validation process and are discussed in Appendix B.

4.0 PROBLEM AREAS

4.1 QUESTIONS ARISING DURING THE PROJECT

The instruction manual supplied by the EPA Project Officer was very complete and made a very successful attempt to deal with all problems which might occur. However, a few questions required answers which were not available from the manual. These questions, along with their answers were documented as they arose, and a copy was given to the Project Officer at the completion of the task, (see Appendix A). The problems raised by the questions should be considered prior to any future revision to the coding procedure manual.

Probably the most serious problem deals with the case where an exhaust gas stream from a single pollution-producing piece of equipment is split into two or more streams, not all of which are sampled. In this situation, there is no correlation between the process rate for the piece of equipment, and the emissions as determined by the test. Either the process rate must be reduced a proportionate amount, or the emissions increased. The problem is what (if any) apportioning factor to use.

Another problem applicable almost exclusively to the EMB data was the lack of process and control equipment efficiency data. Without these two data elements, emission factors cannot be calculated. Some effort should be made to insure that these data are taken during a test, and, equally important, that they are included in the report.

5.0 SUGGESTED FUTURE EFFORT

The two most promising areas for obtaining additional data are probably the individual state control agency offices, and the control equipment manufacturers.

- State Control Agency Offices - Although there are probably less data available from the state offices, they will certainly be easier to obtain. Some states have already expressed an interest in having their data coded into SOTDAT, and it seems unlikely that other states would refuse to make their data available. All states will, however, resent having to supply the manpower necessary for the coding effort.
- Control Equipment Manufacturers - Control equipment manufacturers usually conduct an inlet and outlet stack test whenever a new piece of equipment is delivered, to insure that the guaranteed efficiency is being met. Therefore there is a large amount of test data in existence, but the manufacturers are extremely reluctant to release the data without first making them anonymous. They are afraid of releasing any proprietary information about their customers. However, the great amount of data available, and the usefulness in evaluating control devices may justify the additional time and expense required to obtain it.

6.0 SUMMARY

The effort expended on this project has produced a sizable data bank of SOTDAT data, and the data obtained are a good representation of most types of pollution sources. However, this effort only scratches the surface of what is available. Many other sources of data are available in addition to those discussed in the preceeding section. Some of these are private consulting firms which have conducted tests, industrial trade associations, plants which have either done their own testing or contracted for required tests, and other government agencies which have conducted tests in connection with research and development projects or the preparation of environmental impact statements and/or permit applications. Since the SOTDAT system has the potential to accept a very large amount of additional data, and since there exists a virtually unending supply of data, the data accumulation can continue far into the future, constantly improving and increasing the capabilities and value of the system.

APPENDIX A

QUESTIONS AND ANSWERS CONCERNING SOTDAT CODING

During the initial SOTDAT coding effort, a list of questions was compiled, the answers to which could not be determined from the manual of coding procedures. Those questions are presented here along with the answers supplied by the Project Officer. It is hoped that this will facilitate future coding of source test data by persons unfamiliar with the system.

Q. Are Orsat analyses considered as test results for coding on C cards?

A. No. These data are to be entered in field B 10.

Q. Should control devices listed on D cards be all devices on the piece of equipment or just those indicated in field C 05?

A. Only those in C 05.

Q. If a device control efficiency is unknown, what code should be used?

A. Use the code for a medium efficiency device.

Q. What pollutant code should be used for total gaseous hydrocarbons, since "total" is listed under aliphatic compounds and "gross" is listed under aromatic compounds?

A. Use code 3101.

Q. Are gaseous samples which are taken non-simultaneously with a particulate sample considered part of the same run?

A. If at least one-half of the gaseous sample was taken during the particulate sample, they are considered to be part of the same run. If not, code it on a separate form even though most process stream parameters will be unavailable.

Q. If a traverse point is sampled more than once during a particulate run, how many times is it counted for coding in field B 06?

A. Only once.

Q. If a test is actually performed, and the result is nil or zero (below the detection limit for the method used) should the test be recorded?

A. Yes. Enter the result as zero.

Q. Is the code for particulate caught by a control device to be "total particulate", "filterable particulate", or "condensable particulate"?

A. Use code "A1101" (total particulate), because most particulate devices are designed for controlling both filterable and condensable fractions, and design efficiencies (field D 02) are usually given in terms of total particulate. However, if design efficiencies are given for the other particulate fractions. They should also be entered alongside their respective pollutant codes ("B 1101" for "filterable particulate", and "C 1101" for condensable particulate).

Q. What should be done with data that are either too large or small to "fit" in the field(s) allotted for them on the coding form?

A. Enter in "Comments" (Section E). Fill the appropriate field(s) with nines.

Q. How does one enter a negative pollutant temperature?

A. Leave field C 07 blank, and write the true temperature in "Comments".

Q. If effective duct cross-sectional area is different from the design area (due to negative flow or sediment build up) which should be entered in field B 03?

A. Enter effective area in field B 03 and write the actual area in "Comments".

Q. If a single stack, fed by several gas streams, each containing a different number of control devices is sampled, how many devices are considered to be upstream from the sampling point?

A. The number of devices found in the stream containing the largest number of devices is used. Indicate in comments.

Q. Are operating parameters (field D 05) "operating" or "design" values?

A. Operating. No design data are to be entered in this field.

Q. If the exhaust from a single piece of equipment breaks into two or more separate gas streams, and both streams are tested, what values are entered in fields A 11 and A 12 (activity levels)?

A. None. Leave those fields blank and enter activity levels in "Comments" along with a statement such as:. "This form contains test data from one of three stacks. See form numbers . ..... and . ..... for data on the other stacks".

APPENDIX B

OBSERVATIONS ON THE SOTDAT SYSTEM

During the data validation process, several items (some essential and some not) came to mind concerning ways to improve the SOTDAT System. These were noted at the time, and are discussed in this appendix.

1. The original EPA Project Officer directed that instead of using a great amount of time writing the plant name and address on each form, the name and address be written only on the first form of a series of tests at a plant, and the form number of that first form be written in place of the name and address on subsequent forms. It was supposed to be included in the keypunching instructions that the name and address from the first form be duplicated on the subsequent forms, but the instruction was apparently either not given or misunderstood. Therefore on forms with a form number greater than A 00390, the form number of the first form for a series of tests appears in the output as the name and address on subsequent forms.

2. Test results are coded three per C card. If the computer finds data in the first test results fields it expects to find data in the remaining fields. Therefore fields which are specified as requiring numeric data and are left blank are interpreted as containing illegal characters, and are printed out as ampersands.

3. Related to the previous problem is the problem of how to treat unknown data. All data in the system now were coded assuming (as is the case with NEDS) that unknown data should be left blank, while data with a numeric value of zero should be coded as a zero. Both types of entries are printed out as zero in cases where blanks are legal characters for the field, and ampersands where they are illegal. Some of the fields where they are illegal are; "Control Device Year Installed", "Sampling Location", "Flow Rate", "Flow Rate Units", "Test Method", and "Sampling Location".

4. On almost all records, the NEDS ID data (State, County, AQCR, Plant, and Point) are incomplete. These should be as complete as possible due to the fact that these items are used by the computer, along with run number, to sort and group the stored data. It was decided during the original coding effort that contractor time could be better spent coding data, leaving the NEDS/SOTDAT correlation for NADB personnel. Based on instructions from the project officer, this is the approach which was taken. Determination of the NEDS ID data is a matter of taking the plant's city and state from the SOTDAT form, going to an atlas and looking up that state and city in the index. From the index, the county can be determined. Then the AQCR can be found in AP-102. After the state, county and AQCR NEDS codes have been found, then the NEDS Plant ID can be determined by checking for that plant's name in a listing of NEDS sources. This of course will be successful only if the plant in question has been input into NEDS (not the case for test data from foreign plants, or for data which are to be input into SOTDAT anonymously).

5. Frequently, extra zeroes randomly appear at the beginning of some of the output data fields. Some of these are: "Process Rate" (both capacity and "This Run"), "Test Result," "Cross Section Area," and "Flow Rate".

6. The output would be cleaner and easier to read if some or all of the unused fields (all printed as zero) were suppressed.

7. One digit is often not sufficient to code "Sampling Location". In those cases, to prevent the ampersand when left blank, a nine was coded in that field, and the actual value was written in "Comments". For future efforts, however, it is probably unnecessary to enter the actual value in "Comments" since only the fact that the value is greater than seven is of any significance.

8. Where there is more than one control device entered on a form, the computer drops the pollutant codes for all devices past the first one.

9. During the original coding effort, two forms were inadvertantly coded with form number A 00098. It seems unlikely that either one of them is currently stored in the computer. Additionally, the A 00098 form for the Wood River Power Plant should have 77.32 instead of 30.44 coded in the "Gas Pressure" field. No attempt was made to correct this problem since the proper procedure for correction was unknown.

10. When trace metal sampling results are to be coded in field C 06, the results will often be too small to enter in the field. It is suggested that another Units code be adopted for field C 03 to represent milli-micrograms per cubic meter.